AD-A086 876   ILLINOIS UNIV AT URBANA-CHAMPAIGN DEPT OF PSYCHOLOGY   F/6 12/1
DIFFERENTIALLY WEIGHTING LINEAR MODELS OF BEHAVIOR: AN EMPIRICA--ETC(U)
JUN 80   C K PARSONS, C L HULIN   N00014-75-C-0904
UNCLASSIFIED   TR-80-4   NL

1 OF 1
AD A086876

END
DATE FILMED
8-80
DTIC

LEVEL II

(12)

UNIVERSITY OF ILLINOIS

Studies of Individuals and Groups in Complex Organizations

AD A086876

Department of Psychology
Urbana - Champaign

DIFFERENTIALLY WEIGHTING LINEAR MODELS OF BEHAVIOR:

AN EMPIRICAL COMPARISON OF SIX METHODS

Charles K. Parsons
College of Management
Georgia Institute of Technology

Charles L. Hulin
Department of Psychology
University of Illinois

Technical Report 80-4, June, 1980

Prepared with the support of the Organizational Effectiveness Research Programs, Office of Naval Research, Contract N000-14-75-C-0904, NR 170-802.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 80-4	2. GOVT ACCESSION NO. AD-A086 876	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Differentially weighting linear models of Behavior: An empirical comparison of six methods		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Charles K. Parsons, Charles L. Hulin		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0904
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology, University of Illinois, Champaign, IL 61820		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N-170-802
11. CONTROLLING OFFICE NAME AND ADDRESS Organizational Effectiveness Research Programs, Office of Naval Research, Code 452 Arlington, VA 22217		12. REPORT DATE June, 1980
		13. NUMBER OF PAGES 31
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

14) TR-80-4

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

9) Technical rept.

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

estimating regression weights, ordinary least squares, Bayesian regression, ridge regression, equal weighting, equation stability, differential weighting, equation validity, cross validation, generalized weight equation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Six methods of estimating regression weights for a linear model of behavior were compared in 51 samples of National Guardsmen. Ordinary least squares, Bayesian m-group regression, ridge regression, equal weighting, and two related methods were used. Weights were estimated in one-half of each sample and then applied to data in the other half. Ratios of observations to predictors ranged from 4:1 to 19:1. Cross validation r^2 was used as the index of model or equation stability. Results support ealier findings that least squares weights are relatively unstable in small samples, but do not indicate the superiority of any one other method.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

400139

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (continued)

Future research and implications for using these regression techniques in testing behavioral models are discussed.

Abstract

Six methods of estimating regression weights for a linear model of behavior were compared in 51 samples of National Guardsmen. Ordinary least squares, Bayesian m-group regression, ridge regression, equal weighting, and two related methods were used. Weights were estimated in one-half of each sample and then applied to data in the other half. Ratios of observations to predictors ranged from 4:1 to 19:1. Cross validation r^2 was used as the index of model or equation stability. Results support earlier findings that least squares weights are relatively unstable in small samples, but do not indicate the superiority of any one other method. Future research and implications for using these regression techniques in testing behavioral models are discussed.

Linear equations frequently appear as descriptions of human behavior. A criterion variable is viewed as an additive function of one or more predictor variables. The differential weighting of the standardized predictor variables is often assumed to convey the relative importance of these variables in explaining the criterion. However, when differential weights are generated statistically from a sample of observations, their properties must be understood before strong interpretations can be made. The purpose of this study is to examine empirically several methods of statistically computing weights.

Multiple regression has certainly been the most popular method used to estimate weights in the linear model. Despite warnings about the sampling error of ordinary least squares regression weights (OLS) (Wainer, 1976, 1978) and interpretative ambiguity when predictors are correlated (Darlington, 1968; Johnston, 1972), many researchers continue to base interpretations of results on the relative size of these weights. Johnston (1972) provides three reasons why interpretations must be qualified when the predictor variables are correlated. First, it is difficult to determine the relative influence of various predictors. Second, an investigator may drop a potentially interesting and useful variable from further consideration because its OLS regression weight is not significantly different from zero. Finally, the estimates of the weights become very sensitive to particular samples of data.

When an investigator tests a model in (or generates a model from!) only one sample of data, two interrelated issues must be considered before assuming that OLS provides the "best" weights. First, the sampling error of the weights must temper direct interpretation of the weights. Sampling

error is negatively related to sample size and positively related to the intercorrelations among predictors. Small samples and correlated predictors are quite common in applied psychological research. Second, theoretical interpretations of the OLS equation implies a decision concerning an index of predictor "importance." As Darlington (1968) points out, unless the predictors are mutually uncorrelated, this is a highly ambiguous matter. Beta weights (standardized regression weights) may be only vaguely related to predictor validity (r^2_{yxi}), and heavily influenced by the other variables in the regression equation.

In pure prediction situations, validity of regression equations is often based on the associated multiple correlation.[2] The effect of sampling error on the multiple correlation coefficient can be demonstrated through the application of various shrinkage formulas (Drasgow, Dorans and Tucker, 1979; Schmitt, Coyle, and Raushenburger, 1977; Wherry, 1931) and cross validation (Dunnette, 1966). The latter method may give an underestimate of the long run or population cross validity (Schmidt, 1971) but does provide the researcher with an estimate of shrinkage for specific situations and purposes.

Wainer (1976, 1978) has suggested that the shrinkage associated with OLS is so undesireable that, uncer many circumstances, OLS should be abandoned as a means for establishing weights in the linear model. Rather, he states that the unbiased properties of OLS do not practically reduce the mean squared error (MSE) of prediction over many other biased weighting strategies. In particular, Wainer (1976) indicates that the equal weighting (EW) of all predictor variables results in an equation that is robust to sampling differences and in many cases has a MSE that is only slightly

higher than OLS in the derivation sample.

Some previous research has compared the predictive stability of EW as an alternative weighting scheme to OLS. In early studies, the superiority of EW over OLS was demonstrated in samples of blue collar workers (Trattner, 1963) and freshman engineering students (Lawshe and Shucker, 1959). Dawes and Corrigan (1974) showed that EW provided a linear composite that was more valid than OLS in describing both the process of decision making and its validity. Recently, several investigators (Dorans & Drasgow, 1978; Einhorn & Hogarth, 1975; Schmidt, 1971) have used Monte Carlo procedures to give good indications of when equal weights are superior to OLS weights. When the sample sizes (N) are small relative to the number of predictors (n), [N=25, n=2 (Schmidt, 1971); N=50, n=4 (Einhorn and Hogarth, 1975); N=30, N=60, n=11 (Dorans and Drasgow, 1978)], equal weights are clearly superior to OLS when cross validated in the population or samples of equal size. Einhorn and Hogarth (1975) also noted that as the criterion (or dependent) variable becomes more ambiguous, equal weights are superior to both OLS and rational (clinical) weights. In attempting to recover population regression weights from sample data, which is directly relevant for model testing, Dorans and Drasgow (1978) found that with N=30 and N=60 (with n=11), equal weights had a higher congruence coefficient (Korth and Tucker, 1975) with the population weights than OLS. Another interesting finding by the same authors was that the congruence coefficients for OLS and EW were approximately equal at N=120, though EW had the higher cross validation r^2. They also found that the pattern of EW superiority was less obvious when the communality of the criterion was low in the population and when the prespecified factor model provided only a moderate fit to the population

covariance matrix.

Overall, the Monte Carlo results do point to the efficiency of equal weights in small samples when randomly drawn from multivariate normal populations. The interpretation and generalization of these results to applied settings depends on the tenability of the assumptions. For instance, Schmidt (1971) suggested that in 20% of the data samples in applied psychology, there are significant departures from the assumptions of the regression model; linearity of regression, normality, and homogeneity of conditional variance. Both he and Einhorn and Hogarth (1975) interpreted their results as conservative estimates of the difference between OLS r^2_{cv} and EW r^2_{cv}.

Based on these results, it is tempting to endorse equal weighting as the desired method in many common research situations. However, equal weighting has serious drawbacks as a method for testing behavioral models. Model testing in field settings depends on the ability to assign differential weights to predictor variables based on empirical observation and statistical theory. As Darlington (1978) has noted, equal weights are not based on any such theory.

In addition to equal weights, there have been numerous suggestions for alternatives to OLS estimates of regression weights. This paper considers 2 alternatives, ridge regression and Bayesian m-group regression. (See Dorans & Drasgow, 1978 for a description and Monte Carlo comparison of some others.)

Several authors (Darlington, 1978; Hoerl and Kennard, 1970; Price, 1977; Theobold, 1974) have recommended the use of ridge regression as an approach to reducing the ambiguity in sample estimates of population weights

among correlated predictors. The object is to obtain biased estimators of the true weights that have smaller sampling variances than OLS estimates. Since these authors have all offered descriptions of the method, the present one will be brief. When computing standardized OLS weights from a correlation matrix, the following formula can be used:

$$b = (X'X)^{-1}X'Y$$

where b is the vector of standardized regression weights, X'X is the correlation matrix for the predictors, and X'Y is the vector of correlations between predictors and criterion. In ridge regression, a small positive value (κ) is added to the diagonal elements of the X'X matrix. The formula,

$$b(k) = (X'X + K)^{-1}X'Y,$$

where K is the vector of small positive values that introduces the bias into the regression estimates (b(K)), has the positive effect of reducing sampling variance. The larger the values in K, the smaller the sampling variance. However, as K gets large, the elements of the weight vector b(K) approach zero. Hoerl and Kennard (1970) and Theobold (1974) demonstrated theoretically that there exists a vector of elements in K such that

$$E[(b(K) - B)'(b(K) - B)] \leq E[(b - B)'(b - B)]$$

where b is the vector of OLS sample weights and B is the vector of population OLS weights. The biased estimators, b(K) have better mean squared error properties in the long run than do OLS estimators if the

squared bias is less than or equal to the reduction in sampling variance. There has been no evidence regarding the optimal method for determining K.

Price (1977) applied ridge regression to the problem of predicting job satisfaction from a set of 5 leadership variables that were linearly related in a sample of 30 observations. His results suggest that the method does eliminate negative beta weights when it is likely that the negative OLS estimates are due to sampling error. No evidence was presented supporting the stability of the ridge regression weights.

Bayesian m-group regression, another alternative to least squares regression, has been proposed as a solution to the problem of estimating regression coefficients in possibly different groups when the groups are small (Jackson, Novick, and Thayer, 1971; Novick, Jackson, Thayer, and Cole, 1972). The objective is to improve prediction in the i^{th} group by using information from the other m-1 groups. Prior information for the Bayesian regression is based on a set of assumptions regarding the nature of the regression weights. In the ordinary least squares model, B_i (the vector of weights), is fixed for the i^{th} group. In this Bayesian model, B_i is assumed to be randomly sampled from a multivariate normal distribution. Prior beliefs about the regression weights are then expressed by specifying values for the parameters of the B_i distribution, and the weight to be given to these prior estimates. In applying m-group regression, Novick, et al. (1972) have assumed that only minimal knowledge exists about this distribution of B_i. Therefore, relatively little weight is given to the estimated prior as compared to information from new group-specific data.

These priors, the least squares regression weights within each group, and the number of observations in each group, systematically affect the

estimated weights. The specification of a large population variance of the regression weights indicates that more heterogeneity is believed to exist among the groups. Therefore the m-group regression weights will be similar to the OLS weights. In an extreme case where the groups are believed to be identical, the population variances will be zero and B_i will be identical for all groups. Group size also affects the estimated B_i. For a given set of population weight variances, the number of observations in the i^{th} group is inversely related to the effect of the data from the other m-1 group on the regression within the i^{th} group.

In summary, theory testing and model building in field settings have depended heavily on the use of linear equations. Previous research has failed to provide a psychologically acceptable solution for statistically estimating differential weights in samples of data that are similar to those often encountered in field settings (limited number of observations and correlated predictors).

The commonly used method of least squares regression provides unbiased estimates of population weights, but they are difficult to interpret and have low stability when estimated and applied in small samples. Equal weighting provides a more robust <u>prediction</u> model for most situations but lacks the ability to integrate data and theory. Recently, Bayesian m-group regression and ridge-regression have been offered as alternatives that may provide this integrative ability without sacrificing predictive stability.

However, very little empirical research has actually compared these methods to the more traditional methods in samples that are drawn from field settings. The current study conducts such a comparison across multiple samples of various sizes. Though not providing concrete answers to

questions about relative stability in different sample sizes, the results have a great deal of generalizeability due to the collection of data from human respondents. Several trends can be expected. First, m-group regression uses data from more than one group. This should lead to greater stability than either least squares or ridge regression which use only group-specific data. Second, past evidence has suggested that equal weighting is superior to many other methods in small samples (e.g., Dorans and Drasgow, 1978). Therefore, it is expected to out-perform least squares and ridge regression because the latter two are susceptable to sampling error. Bayesian m-group regression is expected to have greater stability than all other methods because it considers both group-specific data as well as data from other groups. It allows for sample differences while using other data to add stability.

Method

Sample

The data for this study were selected from a total sample of 2079 personnel in 60 units of the Illinois Army National Guard. The initial data are arbitrarily divided into Wave 1 (29 units, N=1169) and Wave 2 (31 units, N=910) corresponding to the time of data collection. The sample sizes of the units range from N=2 to N=80. Other analyses of these same data are reported in Hom and Hulin (Note 1) and Katerburg and Hulin (Note 2).

Each National Guard unit served as a sample of observations. It was felt that the objective differences among the units (Katerburg and Hulin, Note 2) provided a rationale for hypothesizing that all observations were not drawn from the same population (i.e., the institutions are not

identical). The Bayesian m-group regression explicitly allows this prior belief to be incorporated into the estimation of weights for each unit.

In order to maintain a reasonably large sample of units and still retain enough observations to estimate weights in each unit, only units with $N \geq 16$ observations were included to estimate weights for two predictors (n=2). All 29 units were retained from Wave 1 and 22 units were retained from Wave 2. The data from each of these 51 units were randomly divided into a derivation sample and a cross-validation sample. The 2 predictor weights were estimated in the derivation sample and then applied to the data in the cross-validation sample.

Assessment of Predictor Variables

Two sets of 2 predictor variables were chosen, one for the Wave 1 units and one for the Wave 2 units. The first set was chosen such that the two variables were highly correlated. This should drastically increase the sampling error of the least squares weights. The Consideration scale from the Leader Behavior Description Questionnaire (Stogdill & Coons, 1963) and the Satisfaction with Supervisor scale from the Job Descriptive Index, JDI (Smith, Kendall, and Hulin, 1969) were chosen because the median intercorrelation was .72 in the 29 units from Wave 1. These variables were used as predictors in the Wave 1 units.

In the second wave, the objective was to choose 2 predictors that consistently produced the highest multiple correlation with the criterion. The JDI Satisfaction with Work scale and JDI Satisfaction with Coworkers scale were used as predictors in the Wave 2 units (median R = .56).

Assessment of the Criterion Variable

A criterion variable represents a behavior (or set of behaviors) that is theoretically related to the predictor variables. Behavioral intention to reenlist (hereafter referred to as behavioral intention) was chosen because it is related conceptually to withdrawal decisions, and reenlistment data were not available for all respondents. Behavioral intention was measured by a single item with a seven point verbally anchored response scale.

The predictive validity of this item was estimated in a sample of 255 respondents for whom data were available. The correlation between responses to this item and actual reenlistment was .70 (Hom and Hulin, Note 1).

In summary, the Wave 1 model included LBDQ-Consideration and JDI-Supervisor to account for variance in behavioral intention. The Wave 2 model included JDI-Work and JDI-Coworkers to account for variance in behavioral intention.

Estimation of Weights

Weights for the two predictor variables were computed in the derivation samples for six alternate weighting schemes. The first four schemes have been described earlier. These are ordinary least squares (OLS), equal weights (EW), Bayesian m-group regression (BAY), and ridge regression (RIDGE)[3]. Two additional methods serve as baselines for comparative purposes. First, ordinary least squares weights were computed from all data available in the derivation samples. As a model this scheme (OLS-Total) assumes that all units are identical and that the larger number of observations adds to the stability of the weights. The final scheme is

derived from the Bayesian methodology and is called the generalized weight equation (GWE).

Cross-validation involved transforming these weights to raw score weights. Dorans and Drasgow (Note 3) have suggested that this transformation is necessary for all cross validation studies because it does not make the possibly erroneous assumption that the ratios of standard deviation $\frac{SD_Y}{SD_{X_i}}$ will not change from derivation to cross validation. In order to cross validate standardized weights, they should be transformed back to the original metric and applied to the cross validation variance-covariance matrix.

For both GWE and OLS-Total, the scalar for this transformation was $\frac{SD_Y}{SD_{X_i}}$ where SD_Y and SD_{X_i} are based on all of the data in the derivation samples. This is consistent with the assumptions that 1) no prior unique data exists for the cross-validation samples and 2) all units are identical and the greater number of observations gives a more stable estimate of the standard deviation in each unit. For the other 4 schemes SD_Y and SD_{X_i} were based on the data from the derivation sample of each unit individually. These raw score weights were then applied to the cross validation variance-covariance matrix ($C_{(cv)}$) according to the following formula:

$$(1) \qquad r^2_{cv} = \frac{(W_D\; C_{XY(cv)})^2}{S^2_{Y(cv)}(W'_D\; C_{XX(cv)} W_D)}$$

where $C_{XY(cv)}$ is the vector of covariances between the two predictors and the criterion, $C_{XX(cv)}$ is the variance-covariance matrix for the two predictors, W_D is the raw score weight vector computed in the derivation sample and $S^2_{Y(cv)}$ is the variance of the criterion in the cross validation

sample.

Results

Wave 1

Table 1 shows the total sample sizes and standard deviations of JDI-Supervisor (X_1), LBDQ-Consideration (X_2), and behavioral intention (Y) in both the derivation sample and cross validation sample for 29 first wave units. The total samples reported in Table 1 were divided into two samples of approximately equal size for derivation and cross validation of weights. In each unit the small differences between derivation and cross validation samples in standard deviation is best explained as sampling error. On the other hand, the somewhat larger differences across the 29 derivation sample standard deviations (the lowest to highest values are .33 to 2.88 for behavioral intention, 3.73 to 13.56 for JDI Supervisor, and 4.62 to 18.13 for LBDQ-Consideration) support the hypothesis that they were drawn from different populations. As would be expected, there are also large differences across units in the cross validation samples.

Table 2 presents the standardized weights for the two predictor variables from each of the six weighting schemes for all 29 units. These weights are typically used in interpreting regression equations. The choice of 1.0 as a value for the equal weighting is arbitrary. As equal standardized weights, any value could be used for the present study. When the cross validation r^2 is computed, the variance of the predictor composite is irrelevant. The OLS scheme yields the following results. In 12 units, LBDQ receives a negative weight. In 7 units,

JDI-Supervisor receives a larger weight (absolute value) in 14 units and LBDQ-Consideration receives the larger weight in the other 15.

As for the alternate schemes, Bayesian m-group regression eliminates all negative weights for both variables. JDI-Supervisor receives the larger weight in 17 units and the weights are equal in one unit. Ridge regression also eliminates all negative regression weights. In the 14 units that had ridge estimates, LBDQ-Consideration had the larger weight in 6 units and JDI-Supervisor had the larger weight in the other 8 units. The two baseline methods, GWE and OLS-Total, give positive and virtually equal weights to both variables.

The r^2_{cv} for the six weighting schemes in the 29 units appears in Table 3. A negative r^2_{cv} reflects the fact that the r_{cv} was less than 0. In other words, the variables were weighted in the opposite direction. The signs are retained in the Table 3 to illustrate this effect. With OLS, 5 units had a r_{cv} of less than 0. For the other 5 methods, only one unit (No. 5) had a negative r_{cv}.

In averaging these statistics across the 29 units, negative r^2_{cv} was considered a 0. This change increases the average r^2_{cv}, $(\bar{r}^2_{cv})$ for OLS, and yet the method still had the lowest $\bar{r}^2_{cv}$ across the 29 units. Somewhat surprisingly, there was very little difference among the $\bar{r}^2_{cv}$ from the other 4 weighting schemes. No statistics exist to compute the statistical significance of differences among the $\bar{r}^2_{cv}$.

If only the 14 units with RIDGE weights are considered, RIDGE gives the highest r^2_{cv} while OLS gives the lowest. Again it is difficult to assess the stability of these differences.

Wave 2

For this wave, Table 4 presents the total unit sample sizes, standard deviation for the two predictors, JDI-Work (X_1) and JDI-Coworkers (X_2), and behavioral intention (Y) in both derivation and cross validation samples. Again, there is a fairly wide range of standard deviations for all three variables across the 22 units.

Table 5 presents the standardized weights derived from the six methods. By design, the predictors were not as highly correlated as in Wave 1. This may explain the OLS results which are more consistent for JDI-Work (X_1). The only negative weight for this variable appears in unit 7. On the other hand, JDI-Coworkers (X_2) received a negative weight in 12 units. In 9 units, both variables received positive weights. JDI-Work received the higher weight (absolute value) in 18 units.

Bayesian m-group regression eliminated the negative weight for JDI-Work in unit 7. In 11 units, BAY yielded a negative weight for JDI-Coworkers. In 21 of the 22 units, the Bayesian weight for JDI-Work was larger (absolute value) than that for JDI-Coworkers. RIDGE resulted in the negative weight for JDI-Work in unit 7 while yielding a negative weight for JDI-Coworkers in 3 units. In 14 units, JDI-Work received the larger weight. The baseline methods gave a positive weight for JDI-Work and small negative weights for JDI-Coworkers.

Again, these weights were transformed using the same scalers as in Wave 1. The weights were then applied to the cross validation variance-covariance matrix according to (1). Table 6 shows the r^2_{cv} for RIDGE in 17 units and the r^2_{cv} for the other 5 methods in all 22 units. OLS yielded 3 small negative r_{cv}. BAY, RIDGE, GWE, and OLS-Total resulted in

small negative r_{cv} for 2 units.

On the average, OLS and EQUAL did somewhat poorer than the other 3 methods in this wave. Considering only the 17 units where RIDGE weights were computed, the 3 methods that use the data from all of the units (BAY, GWE, and OLS-Total) demonstrated a slightly higher $\bar{r}^2_{cv}$. Surprisingly, EW showed the poorest stability.

DISCUSSION

These weighting schemes were computed as models of human behavior. For the current predictor variables (facets of satisfaction and leadership), positive weighting of both is consistent with previous research and theory that indicates that they are negatively related to organizational withdrawal (Fleishman and Harris, 1962; Smith et al., 1969).

The six computational methods yielded different patterns of weights. First, under the condition of highly correlated predictors, least squares weights predictably vacillated between positive and negative values. The inflated sampling error of these weights was demonstrated in the lower cross validation r^2 associated with them. The Bayesian and ridge regression methods eliminated the negative weights, but only ridge regression showed improved stability. Interpretation of this latter result must be qualified by the non-random selection of the 14 units that may favor ridge regression.

In the second wave of data, the least squares weight for JDI-Work was consistently positive (except for one unit). In contrast to the first wave, this result indicates the increased stability of the sign of an individual weight when the predictors are correlated to a lesser degree. The ridge regression procedure eliminated more negative weights on the second.

This result indicates the increased stability of the sign of an individual weight when the predictors are correlated to a lesser degree. The ridge regression procedure eliminated more negative weights on the second variable, JDI-Coworkers, than did the Bayesian method. However, the baseline methods indicate that a small negative weight may be realistic for the second variable in many samples.

As for the stability of the models, the cross validation results from this study showed mixed support for the hypothesized trends. Surprisingly, the far greater complexity and computing time necessary to estimate the Bayesian weights did not pay off in greater stability in these data. It appears that m-group regression may not be a cost-effective method of estimating weights under conditions similar to those found in the present study. In the first wave of units, the m-group regression stability was only as good as that of equal weighting. In Wave 2, it was equivalent to OLS-Total, indicating that its increase stability over least squares may be due more to its use of more data, than to its methods. OLS-Total is a much less expensive method of obtaining stable and possibly different weights.

Ridge regression did show some promise in the first wave where the predictors were more highly correlated. In the 14 units, its stability exceeded that of the other methods. Its limitation is, of course, its inappropriateness for the other 15 units. The investigator has the option of resorting to least squares weights when ridge constants can not be computed. If this were done in the present data, ridge regression stability would have exceeded that of least squares on the 29 units, but fallen short of that of the other 4 methods. It does have the advantage of needing less data than BAY, GWE, or OLS-Total.

As for the generalizeability of this study, the data were highly characteristic of field studies in general. Small sample sizes, relatively low criterion communality, and limited numbers of samples are very common to the study of behavior in field settings. It would have been gratifying to discover a panacea to the weights estimation problem under these circumstances. Unfortunately, the results of this study suggest that this solution has not yet been achieved. The most that can be concluded from the current study is that least squares provides somewhat poorer estimates than other methods.

One major limitation was that only two predictor variables were included in each equation. As more predictors are added, the estimation and interpretation of weights becomes more complex. When possible, it would be interesting to study the performance of these weighting schemes with more predictors in larger samples. Perhaps under this condition, ridge or Bayesian regression would demonstrate greater stability than equal weighting. Another weakness was the relatively small number of samples, 29 in Wave 1, and 22 in Wave 2. Future research would ideally use more samples of approximately equal sizes in order to obtain better estimates of the differences in stability between estimation methods. Further work should also be conducted on the validity of the estimated weights as opposed to the stability. Since the testing of behavioral models in correlational field studies requires the interpretation of weights, more explicit statements about the accuracy of weighting methods is desired.

For the present, it is argued that because no method demonstrated clear-out superiority over equal weighting, the current state of our research methods does not yet permit us to test differentially weighted

models in small samples. We knew this was true for least squares weights. The present findings suggest that it is also true for several other advocated methods.

Footnotes

1. This research was supported in part by the Office of Naval Research, Contract N000-14-75-C-0904, Charles L. Hulin, principal investigator in part by the Department of Psychology, University of Illinois, and in part by the Illinois National Guard.

2. In order to reduce ambiguity, the following conventions are used for the present study. First, the term, equation validity, will be taken to mean the similarity of the estimated weights to the true population weights. The congruence coefficient (Korth and Tucker, 1975) used by Dorans and Drasgow (1978) is an example of a statistic that indicates equation validity. Second, equation stability will mean the predictive effectiveness of a set of estimated weights in future samples. The indices most commonly used are the shrunken R^2 (e.g., Wherry, 1931) the cross-validated r^2 (r^2_{cv}). The disadvantage of the former is that it is meaningful only for least squares estimates, whereas the latter is appropriate for weights derived from any procedure. The distinction between equation validity and stability is made because the interpretation of the relative size of estimated weights implies that the equation has some validity. However, the estimation of stability through cross-validation may not give much indication about this validity. Darlington (1968) and Wainer (1976) both point out that widely different sets of weights can yield equally stable equations, but of course are not all equally valid. Previous researchers have focused on the stability of the equations. Dorans and Drasgow (1978)

did compute congruency coefficients for different weighting schemes and the results were not identical to those of the r^2_{cv} . Empirical studies (non-Monte Carlo) have no recourse but to use stability as a criterion since population weights are unknown. However studies such as the present one must be interpreted with the distinction between validity and stability in mind.

3. The present authors would like to thank Jerry Isaacs for making the Bayesian m-group regression program available. We would also like to thank Richard Darlington for making the ridge regression program available.

REFERENCE NOTES

1. Hom, P. W. and Hulin, C. L. A comparative examination of four approaches to the prediction of organizational withdrawal. Technical report 78-5, University of Illinois, 1978.

2. Katerburg, R. Jr. and Hulin, C. L. The effects of organizational function on responses: The mediating role of technology and job characteristics. Technical Report 78-4, University of Illinois, 1978.

3. Dorans, N. J. and Drasgow, F. A note on cross validating prediction equations. Unpublished manuscript, University of Illinois, 1979.

REFERENCES

Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.

Darlington, R. B Reduced variance regression. Psychological Bulletin, 1978, 85, 1238-1255.

Dawes, R. M. and Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

Dorans, N. J. and Drasgow, F. Alternative weighting schemes for linear prediction. Organizational Behavior and Human Performance, 1978, 21, 316-345.

Drasgow, F., Dorans, N. J., and Tucker, L. R. Estimators of the squared cross-validity coeffieient: A Monte Carlo investigation. Applied Psychological Measurement, 1979, 3, 387-399.

Dunnette, M. D. Personnel Selection and Placement, Belmont, Calif.: Brooks/Cole Publishing Company, 1966.

Einhorn, H. J. and Hogarth, R. M. Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 1975, 13, 171-192.

Fleishmann, E. A. and Harris, E. F. Patterns of leadership behavior related to employee grievances and turnover. Personnel Psychology, 1962, 15, 43-56.

Hoerl, A. E. and Kennard, R. W. Ridge regression: Biased estimation for non-orthogonal problems. Technometrics, 1970, 12, 55-67.

Jackson, P. H., Novick, M. R., and Thayer, D. T. Estimating regressions in m-groups. British Journal of Mathematical and Statistical Psychology, 1971, 24, 129-153.

Johnston, J. Econometric Methods. New York: McGraw-Hill, 1972.

Korth, B. A. and Tucker, L. R. The distribution of chance congruence coefficients from simulated data. Psychometrika, 1975, 40, 361-372.

Lawshe, C. H. and Shucker, R. E. The relative efficiency of four test weighting methods in multiple prediction. Educational and Psychological Measurement, 1959, 19, 103-114.

Novick, M. R., Jackson, P. H., Thayer, D. T. and Cole, N. S. Estimating multiple regressions in m-groups: A cross validation study. British Journal of Mathematical and Statistical Psychology, 1972, 25, 33-50.

Price, B. Ridge regression: Application to non-experimental data. Psychological Bulletin, 1977, 84, 759-766.

Schmidt, F. L. The relative efficiency of regression and simple unit predictor weights in applied differential psychology. Educational and Psychological Measurement, 1971, 31, 699-714.

Schmitt, N., Coyle, B. W. and Raushenberger, J. A Monte Carlo evaluation of three formula estimates of cross-validated multiple correlation. Psychological Bulletin, 1977, 84, 751-758.

Smith, P. C., Kendall, L. M. and Hulin, C. L. The Measurement of Satisfaction in Work and Retirement. Chicago: Rand McNally, 1969.

Stogdill, R. M. and Coons, A. E. Manual for the Leader Behavior Description Questionnaire-Form XII. Columbus: Ohio State University, Bureau of Business Research, 1963.

Theobald, C. M. Generalization of mean squared error applied to ridge regression. Journal of the Royal Statistical Society, Series B, 1974, 36, 103-106.

Trattner, M. H. Comparison of three methods for assembling aptitude test batteries. Personnel Psychology, 1963, 16, 221-232.

Wainer, H. Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 1976, 83, 213-217.

Wainer, H. On the sensitivity of regression and regressors. Psychological Bulletin, 1978, 85, 267-273.

Wherry, R. J. A new formula for predicting the shrinkage of the coefficient of multiple correlation. Annals of Mathematical Statistics, 1931, 2, 440-457.

TABLE 1

Total Wave 1 Unit Sizes and Standard Deviations of all Variables in Derivation and Cross Validation Samples

Unit	Total N	Derivation SD_Y	Derivation SD_{X_1}	Derivation SD_{X_2}	Cross Validation SD_Y	Cross Validation SD_{X_1}	Cross Validation SD_{X_2}
1	45	2.61	9.95	13.95	2.24	9.98	14.97
2	19	2.30	7.46	14.97	2.79	9.22	14.48
3	49	2.61	7.67	10.63	2.65	5.60	10.19
4	46	2.24	6.89	11.33	2.64	6.54	9.27
5	27	2.44	7.34	11.06	2.75	6.94	9.19
6	37	2.81	9.03	13.43	2.43	8.13	14.93
7	49	2.25	8.69	16.04	2.54	8.66	12.15
8	29	2.62	5.93	11.06	2.61	6.81	11.04
9	32	2.56	6.41	9.89	2.10	4.10	7.99
10	70	2.46	7.40	10.91	2.42	7.07	10.00
11	18	2.37	8.41	10.40	2.55	10.07	10.81
12	22	2.51	10.24	18.13	2.86	10.04	14.29
13	18	2.77	7.52	7.96	3.05	6.95	11.47
14	19	.33	7.01	8.61	2.77	7.55	11.68
15	16	2.88	7.42	12.61	1.69	4.94	8.76
16	50	2.29	6.48	11.42	1.69	5.41	11.32
17	28	2.73	8.63	15.94	2.45	6.31	11.26
18	28	1.95	3.73	6.92	2.35	3.43	5.80
19	34	2.10	7.29	10.68	2.05	4.91	6.36
20	35	2.38	4.93	10.08	2.20	6.24	9.57
21	47	2.34	5.15	7.13	2.54	7.96	10.15
22	23	2.42	13.56	15.12	2.21	6.65	6.85
23	35	2.57	5.78	4.62	2.59	6.12	5.42
24	37	2.70	10.11	16.72	2.46	10.07	17.74
25	34	2.45	7.65	13.05	2.21	7.61	18.15
26	24	2.22	7.59	8.77	2.39	5.60	15.14
27	75	2.45	6.59	8.57	2.86	7.78	10.48
28	36	2.53	7.40	9.29	2.77	7.96	12.06
29	39	2.59	7.14	13.13	2.46	8.84	2.46

TABLE 2

Standardized Regression Weights for Wave 2 Units

	EQUAL		OLS		BAY		RIDGE		GWE		OLS-Total	
Unit	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
1	1.000	1.000	-.272	.278	.080	.091	a		.153	.142	.169	.142
2	1.000	1.000	1.175	-.919	.169	.085			.153	.142	.169	.142
3	1.000	1.000	.305	-.085	.150	.085			.153	.142	.169	.142
4	1.000	1.000	-.089	.379	.123	.207	.029	.066	.153	.142	.169	.142
5	1.000	1.000	.420	.200	.190	.204	.278	.195	.153	.142	.169	.142
6	1.000	1.000	.190	.273	.177	.211	.135	.158	.153	.142	.169	.142
7	1.000	1.000	-.039	.164	.135	.128			.153	.142	.169	.142
8	1.000	1.000	.227	.021	.113	.123			.153	.142	.169	.142
9	1.000	1.000	-.118	.263	.100	.128			.153	.142	.169	.142
10	1.000	1.000	.298	-.138	.157	.044			.153	.142	.169	.142
11	1.000	1.000	-1.044	1.002	.128	.119			.153	.142	.169	.142
12	1.000	1.000	.747	-.036	.298	.188	.551	.043	.153	.142	.169	.142
13	1.000	1.000	-.108	.448	.138	.109			.153	.142	.169	.142
14	1.000	1.000	.231	.336	1.051	.543	.105	.103	.153	.142	.169	.142
15	1.000	1.000	-.514	.281	.083	.044			.153	.142	.169	.142
16	1.000	1.000	-.119	.369	.107	.189	.006	.050	.153	.142	.169	.142
17	1.000	1.000	-.322	.921	.161	.338	.050	.460	.153	.142	.169	.142
18	1.000	1.000	-.145	.529	.101	.146	.040	.114	.153	.142	.169	.142
19	1.000	1.000	.418	.000	.191	.163	.115	.087	.153	.142	.169	.142
20	1.000	1.000	.076	.140	.095	.135			.153	.142	.169	.142
21	1.000	1.000	.365	-.113	.132	.073	.031	.046	.153	.142	.169	.142
22	1.000	1.000	.224	.156	.246	.206			.153	.142	.169	.142
23	1.000	1.000	.103	.690	.130	.111	.110	.600	.153	.142	.169	.142
24	1.000	1.000	.588	-.149	.217	.167	.199	.105	.153	.142	.169	.142
25	1.000	1.000	-.437	.468	.103	.090			.153	.142	.169	.142
26	1.000	1.000	.212	.178	.154	.154			.153	.142	.169	.142
27	1.000	1.000	-.008	.056	.091	.052			.153	.142	.169	.142
28	1.000	1.000	.649	-.209	.202	.114	.301	.022	.153	.142	.169	.142
29	1.000	1.000	.014	.654	.165	.314	.121	.482	.153	.142	.169	.142

[a] The blanks for RIDGE indicate that no values of K, the vector of ridge constants, would improve OLS. This convention holds for Tables 2, 3, 5 and 6.

TABLE 3

Cross Validation r^2 for Six Weighting Schemes in Wave 1

Unit	EQUAL	OLS	BAY	RIDGE	GWE	OLS-Total
1	.040	-.044	.038		.041	.044
2	.054	.070	.058		.052	.053
3	.213	.0-2	.184		.208	.199
4	.221	.179	.224	.224	.221	.219
5	-.002	.003	-.003	.000	-.002	-.000
6	.056	.053	.055	.055	.057	.058
7	.179	.151	.178		.182	.179
8	.159	.113	.161		.164	.160
9	.207	-.010	.168		.208	.236
10	.239	.088	.208		.237	.233
11	.415	-.154	.415		.415	.415
12	.144	.203	.158	.192	.139	.145
13	.193	.236	.188		.184	.180
14	.575	.585	.547	.574	.566	.559
15	.012	-.014	.018		.013	.014
16	.014	.019	.018	.019	.015	.014
17	.170	.221	.192	.212	.176	.170
18	.105	.014	.096	.077	.101	.105
19	.040	.054	.042	.043	.041	.043
20	.140	.160	.152		.150	.144
21	.157	.017	.126	.176	.151	.142
22	.056	.045	.050		.046	.042
23	.121	.130	.118	.130	.104	.100
24	.006	-.023	.004	.001	.007	.005
25	.083	.092	.079		.085	.081
26	.000	.000	.000		.000	.000
27	.340	.116	.376		.357	.364
28	.039	.079	.049	.068	.043	.046
29	.118	.105	.116	.112	.117	.118
Mean (29 units)	.141	.095	.138		.141	.140
Mean (14 units)	.126	.119	.125	.135	.124	.123

TABLE 4

Total Wave 2 Unit Sizes and Standard Deviations of all Variables in Derivation and Cross Validation Samples

		Derivation			Cross Validation		
Unit	Total N	SD_Y	SD_{X_1}	SD_{X_2}	SD_Y	SD_{X_1}	SD_{X_2}
1	22	2.453	14.051	7.321	2.541	14.795	5.947
2	30	2.274	11.836	10.369	2.251	11.111	9.812
3	68	2.600	10.086	9.770	2.722	11.197	8.285
4	48	2.302	10.287	6.022	2.371	11.489	8.715
5	60	2.315	13.393	9.972	2.318	12.905	7.021
6	51	1.925	8.912	5.923	1.742	10.498	6.211
7	53	1.587	5.622	4.039	1.272	6.276	3.979
8	35	2.351	13.122	9.162	2.209	8.852	9.342
9	29	2.349	9.221	11.169	2.295	12.252	9.395
10	20	2.003	12.985	9.661	2.068	15.094	4.590
11	41	2.408	10.238	9.370	2.250	10.929	8.555
12	25	2.065	11.474	9.209	2.293	8.861	6.640
13	18	1.394	15.951	14.560	2.455	10.620	9.015
14	17	2.866	13.750	13.328	2.682	8.847	3.317
15	28	2.701	13.392	8.976	2.240	14.872	9.843
16	20	1.767	13.267	7.704	2.283	11.559	3.653
17	26	2.362	11.414	5.540	1.377	7.967	2.309
18	33	2.380	9.879	9.059	1.954	10.241	7.064
19	25	2.000	13.198	5.299	2.410	9.721	4.997
20	18	2.550	13.546	9.329	2.472	10.161	4.512
21	44	2.334	11.591	8.423	2.476	12.261	6.448
22	29	2.336	11.494	9.758	2.274	11.836	10.369

TABLE 5

Standardized Regression Weights for Wave 2 Units

Unit	EQUAL X_1	EQUAL X_2	OLS X_1	OLS X_2	BAY X_1	BAY X_2	RIDGE X_1	RIDGE X_2	GWE X_1	GWE X_2	OLS-Total X_1	OLS-Total X_2
1	1.000	1.000	.985	-.507	.493	-.036	.676	.-257	.412	-.016	.429	-.012
2	1.000	1.000	.349	-.046	.422	-.027			.412	-.016	.429	-.012
3	1.000	1.000	.427	.139	.334	.135	.349	.147	.412	-.016	.429	-.012
4	1.000	1.000	.626	-.570	.371	-.235	.447	-.403	.412	-.016	.429	-.012
5	1.000	1.000	.335	.207	.457	.108	.255	.177	.412	-.016	.429	-.012
6	1.000	1.000	.162	.301	.384	.163	.120	.182	.412	-.016	.429	-.012
7	1.000	1.000	-.014	.094	.298	.043			.412	-.016	.429	-.012
8	1.000	1.000	.212	-.132	.435	-.207			.412	-.016	.429	-.012
9	1.000	1.000	.130	-.732	.302	-.613	-.102	-.395	.412	-.016	.429	-.012
10	1.000	1.000	.538	.212	.538	.121	.266	.121	.412	-.016	.429	-.012
11	1.000	1.000	.467	-.062	.360	-.016	.210	.043	.412	-.016	.429	-.012
12	1.000	1.000	.544	-.415	.456	-.205	.300	.219	.412	-.016	.429	-.012
13	1.000	1.000	.744	-.365	.949	.052	.526	.230	.412	-.016	.429	-.012
14	1.000	1.000	.739	-.214	.408	-.037	.273	.009	.412	-.016	.429	-.012
15	1.000	1.000	.278	-.163	.392	-.133			.412	-.016	.429	-.012
16	1.000	1.000	.083	.078	.578	-.092			.412	-.016	.429	-.012
17	1.000	1.000	.831	-.441	.406	-.009	.164	.040	.412	-.016	.429	-.012
18	1.000	1.000	.805	-.118	.365	.091	.652	-.023	.412	-.016	.429	-.012
19	1.000	1.000	.469	.289	.541	.042	.238	.146	.412	-.016	.429	-.012
20	1.000	1.000	.441	.289	.452	.139	.303	.245	.412	-.016	.429	-.012
21	1.000	1.000	.417	.079	.407	.032	.204	.032	.412	-.016	.429	-.012
22	1.000	1.000	.039	.710	.394	.330	.093	.572	.412	-.016	.429	-.012

TABLE 6

Cross Validation r^2 for Six Weighting Schemes in Wave 2

Unit	EQUAL	OLS	BAY	RIDGE	GWE	OLS-Total
1	.092	.303	.289	.305	.284	.284
2	.417	.110	.138		.147	.152
3	.362	.311	.321	.323	.241	.244
4	.014	.000	.001	.000	.011	.012
5	.159	.188	.210	.182	.212	.212
6	.061	.025	.117	.036	.187	.185
7	.232	.071	.252		.242	.243
8	.000	.150	.150		.066	.064
9	.187	-.020	.005	.007	.172	.173
10	.000	-.001	-.002	-.001	-.005	-.005
11	.138	.163	.169	.176	-.169	.170
12	.020	.391	.380	.057	.293	.290
13	.294	.100	.243	.297	.220	.223
14	.452	.527	.528	.527	.528	.528
15	.086	.056	.081		.102	.103
16	.007	.007	.000		.000	.000
17	.000	-.015	-.006	-.002	-.006	-.006
18	.014	.034	.024	.031	.031	.031
19	.171	.192	.245	.192	.258	.257
20	.534	.470	.402	.515	.288	.291
21	.289	.393	.420	.401	.447	.445
22	.050	.001	.059	.003	.117	.116
Mean (22 units)	.160	.160	.183		.183	.183
Mean (17 units)	.164	.183	.201	.180	.203	.204

ATE

LMED

-8